



Bilan 2011 de la plateforme ISIDORE et perspectives 2012-2015

Stéphane Pouyllau, Jean-Luc Minel, Shadia Kilouchi, Laurent Capelli

► To cite this version:

Stéphane Pouyllau, Jean-Luc Minel, Shadia Kilouchi, Laurent Capelli. Bilan 2011 de la plateforme ISIDORE et perspectives 2012-2015. Comité de pilotage du TGE Adonis. 2012, pp.1-23. sic_00690558v2

HAL Id: sic_00690558

https://archivesic.ccsd.cnrs.fr/sic_00690558v2

Submitted on 3 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License



Bilan 2011
Perspectives 2012-2015
Version 1.1 – 12 mars 2012

Ce rapport a été rédigé par Jean-Luc Minel (TGE Adonis / Université Paris Ouest la Défense) et Stéphane Pouyllau (TGE Adonis / CNRS), responsables du développement d'ISIDORE avec les contributions de Shadia Kilouchi (TGE Adonis / CNRS) et de Laurent Capelli (Centre pour la communication scientifique directe / CNRS). Les auteurs remercient Ariane Allet (TGE Adonis / CNRS) pour la relecture.

Table des matières

Préambule.....	3
Analyses.....	5
Répartition disciplinaire des ressources numériques proposées par ISIDORE.....	5
Éléments et indicateurs généraux.....	6
Tendance générale.....	7
Les visites uniques.....	7
Fidélisation des visiteurs.....	8
Analyse de la durée de la visite.....	8
Analyse des entrées dans ISIDORE.....	8
Analyse des sorties d'ISIDORE.....	9
Nombre de requêtes adressées au moteur de recherche.....	9
Prospective.....	10
Bref état de lieux.....	10
Améliorer l'utilisation des services d'ISIDORE.....	11
ISIDORE comme outil d'accès aux corpus produits par les scientifiques	11
Exploiter les historiques des consultations pour proposer des recherches sur profils.....	12
Proposer un service permettant de faire de la veille scientifique.....	12
Contribuer à la construction d'un vocabulaire scientifique pour le web de données	12
Améliorer la qualité des données moissonnées.....	12
Analyser les usages des étudiants.....	13
Intégrer des données scientifiques multilingues.....	14

Préambule

Ce rapport vise à analyser les utilisations de la plate-forme ISIDORE entre le 1er mai et le 15 octobre 2011 à partir des rapports fournis par les logiciels *Google Analytics* et *Piwik*¹. Dans un second temps, et en nous appuyant sur les observations et les enseignements qu'il est possible d'en tirer, nous proposerons des pistes de réflexion en vue d'élaborer une feuille de route de pilotage, qui dépasse le simple cadre de l'évolution technique de la plate-forme ISIDORE, pour les quatre années à venir. Il s'agit de réfléchir à la place et au rôle d'ISIDORE dans le paysage de la recherche en SHS afin d'anticiper au mieux les évolutions – toujours nécessaires pour un moteur de recherche – que cet outil d'accès unifié aux données devra recevoir dans les années qui viennent.

Nous indiquons au lecteur qu'en 2011, une convention fixant les conditions de la maîtrise d'ouvrage (MOA) et de la maîtrise d'œuvre (MOE) a été signée entre le TGE Adonis et le Centre pour la communication scientifique directe². La partie « bilan » s'inscrit donc dans ce contexte.

Nous plaçons également la seconde partie dans le cadre de l'évolution des institutions pouvant fournir des données numériques. Elles sont en cours d'évolution et se structurent (TGIR Corpus, ANR Corpus, Equipex, Labex). Notons qu'elles ont fortement changées depuis l'ouverture d'ISIDORE en décembre 2010.

Il convient, avant de présenter les indicateurs, de rappeler les différents événements qui ont ponctué la mise en exploitation de la plate-forme ISIDORE.

- **Décembre 2010** : ouverture de la plate-forme ISIDORE, en version de test (Béta), avec 37 collections contenant 800 000 ressources (une ressource correspondant à une notice enrichie avec l'indexation produite par ISIDORE) ;
- **Février 2011** : passage à la version 1 d'ISIDORE ;
- **Avril 2011** : communiqué de presse³ du service de la communication du CNRS et ouverture officielle d'ISIDORE ;
- **Juillet 2011** : recette des logiciels commandés dans le cadre du marché public passé en octobre 2009 pour la réalisation du cœur d'ISIDORE (moteur de recherche, chaînes de traitement des données, etc.). Extension à plus de 1 100 000 ressources, avec notamment le moissonnage de la plate-forme hypotheses.org et le passage au moissonnage incrémental ;
- **Septembre 2011** : extension des collections (49) contenant 1 307 730 ressources issues de plus de 1 000 sources de données (entrepôts de données, bases de données, revues, etc.) ; Mise en service de l'API⁴ et du *Sparql endpoint* pour la réutilisation des métadonnées enrichies⁵.
- **Novembre 2011** : signature de la convention MOA/MOE avec le CCSD.

Remarques générales sur le fonctionnement d'ISIDORE depuis décembre 2010 :

- Les interfaces de consultation n'ont pas subi de transformations notables depuis la mise en service ;
- La plate-forme a été opérationnelle 24h/24 sans aucune interruption ;
- Des présentations de ISIDORE ont été réalisées dans plusieurs maisons des sciences de

1 Google Analytics est une plate-forme extérieure au projet, Piwik est l'outil interne de statistiques.

2 Le Centre pour la communication scientifique directe (CCSD) assure l'exploitation d'ISIDORE depuis le 4 avril 2011 après avoir participé activement depuis 2009 à sa conception et à son développement.

3 Nous renvoyons le lecteur à la page : <http://www.tge-adonis.fr/service/isidore> [d.c. 2011/12/21].

4 L'API permet d'utiliser toutes les fonctionnalités du moteur de recherche d'ISIDORE dans des applications tierces (sites web de laboratoires, de bibliothèques de recherche, etc.). L'API d'ISIDORE est proposée sous la forme d'un démonstrateur et d'une documentation électronique à l'adresse : <http://rechercheisidore.fr/api/> [d.c. 2011/12/12].

5 Le *Sparql endpoint* est une interface d'interrogation des métadonnées enrichies d'ISIDORE utilisant le langage d'interrogation normalisé Sparql proposé par le W3C dans le cadre du web sémantique. Le *Sparql endpoint* d'ISIDORE est disponible à l'adresse : <http://rechercheisidore.fr/sparql> [d.c. 2011/12/13].

l'homme ou MSH (Dijon, Lyon, Nanterre, Besançon), dans différents laboratoires (UMR, EA), dans différents centres de documentation et bibliothèque de recherche, au salon I-Expo (Paris, avril 2011), à la conférence de l'ADBU (Vannes, septembre 2011), au salon Documation 2011 et SEARCH 2011 (Paris, mars 2011), par diverses interventions audiovisuelles (France Culture, BFM Business) et par visioconférence avec l'Université de Montréal (proposant depuis des données dans ISIDORE). Par ailleurs, ISIDORE a été relayé par la presse spécialisée (01 Business & Technologies⁶ ; Liberation.fr⁷ ; Documentaliste, Science de l'information⁸ et dans de très nombreux blog scientifiques.

6 Voir : http://www.tge-adonis.fr/sites/default/files/exp_cnrs-1.pdf (avec l'aimable autorisation de 01 B&T) [d.c. 2011/12/14].

7 Voir : <http://sciences.blogs.liberation.fr/home/2011/04/le-cnrs-lance-isidore-les-shs-sur-le-web.html> [d.c. 2011/12/15].

8 A paraître en janvier 2011, cf. <http://www.adbs.fr> [d.c. 2011/12/21].

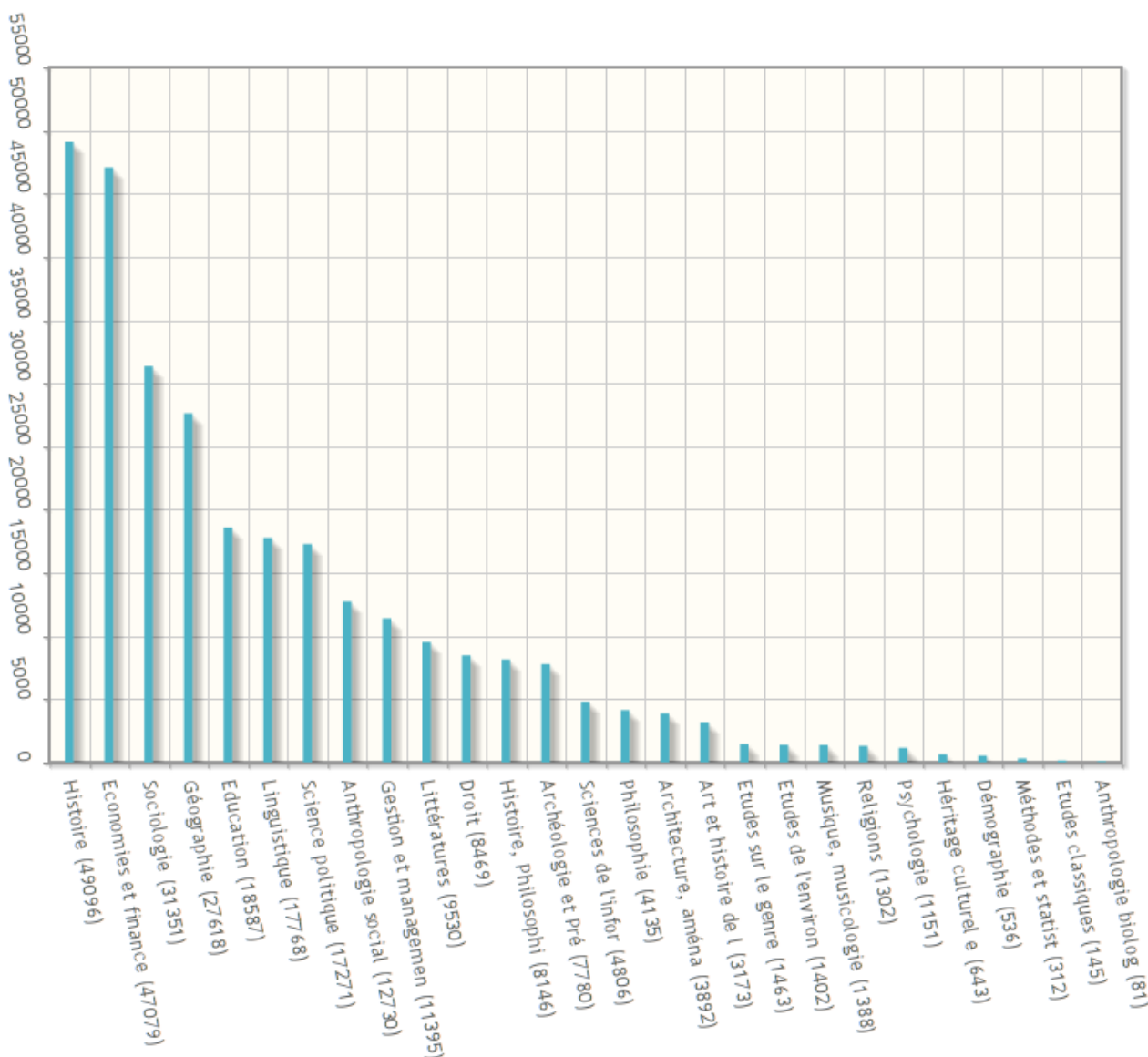
Analyses

Nous analysons ci-dessous les différents indicateurs pour la période du 1er mai au 30 octobre 2011⁹. Cette période d'étude a été choisie en raison du fait qu'elle couvre à la fois la période post-ouverture de la V1 et la première rentrée universitaire pour ISIDORE.

Répartition disciplinaire des ressources numériques proposées par ISIDORE

ISIDORE catégorise les ressources numériques à l'aide des disciplines de HALSHS alignées sur celle de Calenda.org. Cela est réalisé automatiquement par comparaison sémantique à l'aide d'un corpus de référence. Nous avons choisi pour cela les articles déposés dans HALSHS en raison du fait que c'est un ensemble librement accessible d'une part et que les articles sont catégorisés à la main par les déposants d'autre part. ISIDORE catégorise alors automatiquement en fondant son analyse sur les mots du titre, les mots-clés, les mots du résumé des métadonnées. Dans ce cadre, toutes les ressources ne peuvent pas être catégorisées.

Cependant, il est possible d'analyser la répartition disciplinaire des ressources catégorisées :



⁹ Dans cette version nous avons inclus, à titre d'indication les chiffres de novembre 2011 à février 2012.

Le graphique ci-dessus représente la répartition des ressources d'ISIDORE par disciplines scientifiques de HALSHS au 29 février 2012 (il est possible de consulter ce graphique, actualisé en temps réel à l'adresse : <http://www.rechercheisidore.fr/outils/isimashups>).

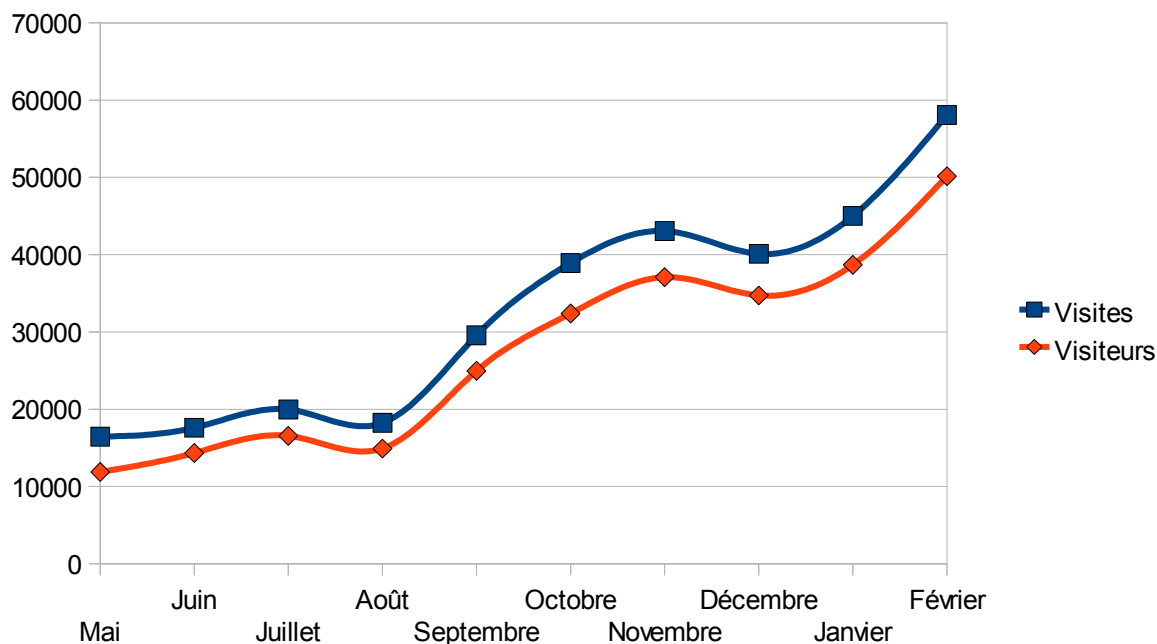
Éléments et indicateurs généraux

	Mai	Juin	Juillet	Août	Septembre	Octobre
Visites uniques	16 452	17 599	19 963	18 224	29 566	38 492
Visiteurs uniques	11 882	14 333	16 549	14 913	24 962	32 385
Actions ¹⁰	55096	44344	46642	43330	79955	103487
Pourcentage de visiteurs connus	NC	NC	12,70%	18,40%	16,63%	20,30%
Taux de rebond	54,00%	56,00%	64,00%	59,00%	54,00%	53,00%
Nombre d'actions par visites	3,3	2,5	2,3	2,4	2,7	2,7
Nombre de pages visitées	NC	NC	20307	38177	70091	89220
Durée moyenne de la visite	2mn53	2mn03	1mn44	1mn49	2mn07	2mn06
Les trois premiers sites entrants	Netvibes.com biblioshs tge-adonis	Google.com Netvibes. biblioshs	Google.com biblioshs tge-adonis	Google.fr biblioshs google.com	Google.fr clg.cq.ca Netvibes	Google.fr Google.com Netvibes
Les trois premiers pays consultants	France USA Canada	France USA Belgique	France USA Belgique	France USA Belgique	France Canada USA	France Canada USA
Les trois premiers sites sortants	Persée Hal-SHS Cairn	Persée Hal-SHS Cairn	Hal-SHS Persée Gallica	Cairn Gallica Hal-SHS	Cairn Gallica Hal-SHS	Cairn Hal-SHS Persée

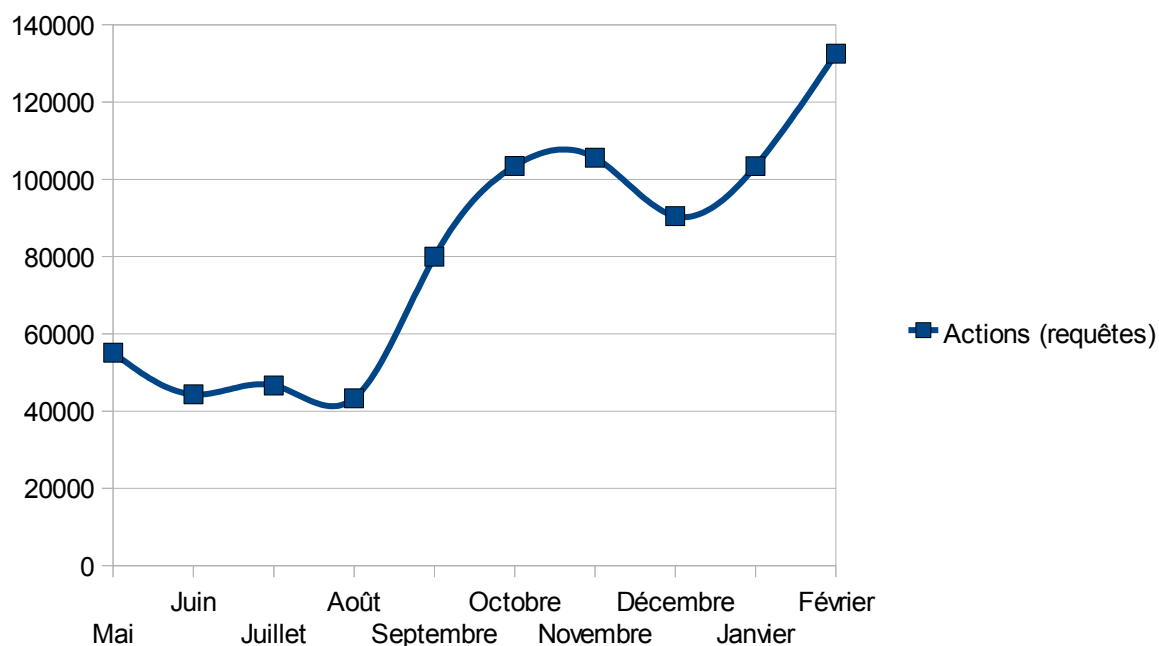
¹⁰ Une action correspond à une requête adressée au moteur.

Tendance générale

La courbe ci-dessous donne la tendance générale des visites uniques entre le 1er mai et le 29 février 2011. Le nombre de visites uniques augmente régulièrement, voir en annexe la dernière synthèse mensuelle.



Graphique 1 : Tendance générale des visites uniques et des visiteurs uniques entre mai 2011 et février 2012



Graphique 2 : Tendance générale des actions (requêtes adressées au moteur de recherche) entre mai 2011 et février 2012

Les visites uniques

Une visite est considérée comme unique si le temps entre deux consultations dépasse 30mn. Ce chiffre est néanmoins à prendre avec précaution puisqu'il est dépendant de l'utilisation des

« cookies » des navigateurs web qui peuvent être effacés par l'utilisateur. En 6 mois le nombre de visites a augmenté de 137%.

Fidélisation des visiteurs

Le pourcentage de visiteurs connus est d'environ 16%. Compte tenu de la forte croissance des consultations ce chiffre peut être considéré comme bon.

Analyse de la durée de la visite

La durée moyenne d'une visite est de 2mn : ce qui correspond au temps nécessaire pour formuler une requête et accéder à l'article recherché soit si la requête est couronnée de succès, soit pour quitter le site. En revanche, cela signifie que ces utilisateurs utilisent peu ou pas les facettes « naviguer ». Ces facettes, présentes sur la une d'ISIDORE, offrent la possibilité de découvrir le contenu d'ISIDORE sans formuler de requête. Ces mêmes facettes sont aussi utilisées pour affiner la recherche (voir annexe 1 et 2) et dans l'onglet « rebondir », or une analyse plus fine des temps de consultation montre qu'environ 10% des utilisateurs ont un temps de consultation compris entre 3 et 30 mn. On peut en conclure que ces utilisateurs utilisent bien ces facettes d'affinage ainsi que les autres fonctionnalités comme « rebondir » ou le nuage de tags proposant les enrichissements d'ISIDORE.

En vue d'améliorer l'utilisation d'ISIDORE, il serait intéressant d'identifier un panel d'utilisateurs afin de mieux comprendre l'usage fait des facettes.

Analyse des entrées dans ISIDORE

Le moteur de recherche Google.fr est le premier site par lequel les utilisateurs accèdent à ISIDORE, le site biblioshs.inist.fr, d'abord deuxième, est maintenant relégué en quatrième position.

On peut faire l'hypothèse suivante : les utilisateurs ne sont plus en majorité des membres de la communauté scientifique de l'enseignement supérieur et de la recherche, en effet, ceux-ci connaissent et utilisent plutôt biblioshs.inist.fr comme porte d'entrée et cela en complément à d'autres accès (Google Scholar par exemple). Par ailleurs, en octobre 2011, l'arrivée de visiteurs via des sites web canadiens et québécois (exemple : le collège Lionel-Groulx www.clg.qc.ca), faisant suite à plusieurs démarches¹¹ et télé-démonstration d'ISIDORE outre-atlantique¹² dont une dans le cadre des cours de l'École d'architecture de l'information (Univ. de Montréal – ENS Lyon), montre tout l'intérêt de développer une politique de communication pour mieux faire connaître ISIDORE. Depuis le mois de décembre, le portail web du service commun de documentation de l'Université de Bordeaux 3 (SHS) propose en complément à l'interrogation de ses propres catalogues une extension de recherche vers ISIDORE¹³ et d'autres intégrations de ce type sont en cours. Cette initiative valide l'idée de proposer ISIDORE sous plusieurs formes afin d'en diversifier l'utilisation (via l'API ou le SPARQL endpoint). Il s'agit d'amplifier ce mouvement.

Les deux pays les plus consultants sur la période des six mois sont tout d'abord la France puis les Etats-Unis. Ces derniers ne représentent que 5% des accès français. Depuis Août 2011, pour la même raison que celle évoquée ci-dessus, le Canada est devenu le deuxième pays consultant, avec un taux égal à 10% des accès français.

Analyse des sorties d'ISIDORE

Depuis le mois d'août 2011, la plate-forme CAIRN¹⁴ est devenue le premier site de sortie

11 Mission de Sophie David, en charge des questions internationales au TGE Adonis en mai 2011.

12 Réalisées par Stéphane Pouyllau, voir : <http://archinfo.umontreal.ca/ee2011/conferences/#interfaces> [d.c. 2011/12/12].

13 Voir : <http://scd.u-bordeaux3.fr/babordplus/> [d.c. 2011/12/12] et annexe 6.

14 Voir : <http://www.cairn.fr> [d.c. 2011/12/12].

devant Gallica¹⁵ puis HAL-SHS¹⁶. Il conviendrait de réfléchir aux raisons de la popularité d'un site proposant un accès payant.

En effet, l'accès à CAIRN étant payant pour certaines revues et articles, cela veut-il dire que la majorité des personnels de l'enseignement supérieur ont accès à cette plate-forme dans le cadre des accès négociés par les SCD (via Couperin) ou bien par le site biblioshs.inist.fr (via l'INIST-CNRS) ?

Un autre indicateur pourrait étayer cette idée : en un an d'exploitation d'ISIDORE, aucun utilisateur n'a pris contact avec le TGE, pour indiquer qu'il n'avait pas pu avoir accès à un article de CAIRN, même ceux payants.

Nombre de requêtes adressées au moteur de recherche

ISIDORE comprend un moteur de recherche¹⁷ pour permettre la recherche d'informations dans les notices qui ont été préalablement collectées puis enrichies par les chaînes de traitement que nous avons mis en place. Le nombre de requêtes adressées au moteur – c'est-à-dire le fait qu'un utilisateur lance une recherche – est un indicateur important dans les outils d'accès à l'information. Dans notre plage d'étude, le nombre de requêtes adressé au moteur de recherche est en forte augmentation depuis le mois de septembre. Elles ont presque doublé entre août et octobre. Cela semble plus dû à l'augmentation de la fréquentation en visiteurs qu'au temps passé sur le site qui lui reste stable ainsi que le nombre d'actions par visites.

15 Voir : <http://gallica.bnf.fr> [d.c. 2011/12/13].

16 Voir : <http://halshs.archives-ouvertes.fr> [d.c. 2011/12/12].

17 Depuis son ouverture en 2010, ISIDORE utilise le moteur de recherche AFS d'Antidot pour rechercher de l'information dans les notices enrichies.

Prospective

Bref état des lieux

Le développement des technologies numériques a transformé les conditions de production et de diffusion des connaissances fondamentales produites par la recherche. L'édition, et plus particulièrement l'édition scientifique, a été profondément bouleversée dans ces dix dernières années [Dacos & Mounier 2011]. Tout d'abord, par le mouvement des Archives ouvertes (Open Archive Initiative) [Ginsparg 1996] ponctué par différents jalons, Déclaration de Budapest¹⁸ 2001 puis Déclaration de Berlin¹⁹ de 2003 qui voit les grandes institutions scientifiques (CNRS et INRIA pour la France) appuyer l'initiative.

Ensuite, par la publication en libre accès, par lequel un grand nombre de revues (« *Open access journals* ») donnent directement accès à leurs récentes publications. Dans le fonctionnement de la recherche, les transformations les plus directes ont trait à la disponibilité accrue des données et des connaissances produites par autrui [Bo 2007], qui permettent la mesure des phénomènes sociaux et le croisement des connaissances. Parallèlement, les chercheurs se sont dotés d'outils pouvant traiter de vastes corpus [Dargentas et alii, 2012]. La recherche est également transformée sur le plan organisationnel : L'évaluation des chercheurs est transformée par la bibliométrie [Filliatreau 2008], par la pratique de déclaration en ligne des activités et l'accessibilité nouvelle des rapports et des travaux aux évaluateurs, par la transformation du financement de la recherche « sur projet » [Becker 2009] et par le management de la recherche, rendus possible par le développement des technologies numériques et des modes de circulation rapides et extensifs de l'information.

Le mouvement de l'ouverture des données publiques est encore peu centré sur les données de la recherche et plus particulièrement sur les archives des scientifiques. Mais les premières initiatives sont là, ISIDORE est en France la plus importante, d'autres projets sont en cours.

Mais depuis 2005, avec la mise en place des Centres de ressources numériques du CNRS, mais aussi avec les appels « Corpus » de l'Agence nationale de la recherche, la mise à disposition de grand corpus s'est accélérée et amplifiée. De nombreux réservoirs de données sont maintenant disponibles et la plupart sont interopérables selon des standards internationaux²⁰ (tel qu'OAI-PMH, ou selon la structuration de l'information en RDF). La montée en puissance de la TGIR Corpus en 2011 augmentera, dans les mois et années qui viennent, la présence de corpus de documents numériques d'archives ou de fonds dans ISIDORE. L'existence de plate-formes d'hébergement de données dans les Maisons des sciences de l'homme (Aix, Bordeaux, Caen, Dijon, Lyon, Nanterre par exemple) déjà compatibles avec ISIDORE facilitera le signalement de ces corpus dans notre plate-forme. Il s'agit maintenant de dimensionner ISIDORE pour qu'il puisse croître et se transformer au fil des besoins en matière d'accès et de traitements que la communauté scientifique définira. En effet, l'une des clefs de la bonne évolution d'un moteur de recherche et de se transformer régulièrement et d'anticiper les changements d'usages.

Enfin, il faut mentionner les problèmes soulevés en matière de droit et en particulier par l'évolution de la propriété intellectuelle à l'heure du numérique [Maurel 2008].

Au regard de ces éléments, ISIDORE est aujourd'hui positionné et dimensionné pour l'accès aux données et documents en libre accès, au sens de l'*open access*. Il est centré sur l'accès aux données de corpus, d'archives, d'événements scientifiques et de publications : c'est une spécificité très appréciée par nos utilisateurs. Il est donc complémentaire aux portails d'accès pour les documents payants²¹ (données bibliographiques, revues, abonnements que l'on trouve par exemple

18 Voir : <http://www.soros.org/openaccess> [d.c. 2011/12/19].

19 Voir : <http://oa.mpg.de/berlin-prozess/berliner-erklarung/> [d.c. 2011/12/18].

20 Nous invitons le lecteur à consulter l'annuaire des sources d'ISIDORE : <http://rechercheisidore.fr/annuaire> [d.c. 2011/12/19].

21 Des réflexions sont en cours avec la DIST du CNRS, l'INIST et l'INSHS afin de faciliter les échanges entre ISIDORE et des portails tel que biblioshs. Il ne s'agit pas de fondre des portails ayant des logiques très différentes

dans biblioshs). Nous rappelons qu'il n'existait pas en France, de point d'accès unifié pour les données SHS en libre accès.

Améliorer l'utilisation des services d'ISIDORE

ISIDORE est accessible depuis un site web : rechercheisidore.fr, mais il est possible d'utiliser ISIDORE différemment : via un logiciel extérieur par exemple ou bien directement depuis la fonction rechercher de son ordinateur²². Cela est possible par l'utilisation de l'API²³. Le Conseil scientifique du TGE Adonis, dans sa session du 16 janvier 2012, a souligné l'importance de développer l'utilisation de cette API dans les interfaces de recherche des bibliothèques universitaires en particulier. Ainsi 3 actions pourraient être lancées :

- Communications auprès des Services communs de documentation (SCD) sur les possibilités de l'API ;
- Mise en place d'un atelier/séminaire sur l'API pour les développeurs/webmasters et les éditeurs de logiciels du domaine afin d'encourager l'appropriation l'API et le développement des possibilités d'interroger ISIDORE depuis des sites ou logiciels tiers ;
- Mise en place de démonstrateurs ;

ISIDORE comme outil d'accès aux corpus produits par les scientifiques

La montée en puissances des projets ANR Corpus, des corpus produits dans les Maisons des sciences de l'Homme ou par les centres de ressources numériques de l'INSHS, des TGIR et notamment de la TGIR Corpus pose la question de l'élargissement massif aux données de corpus produits par les scientifiques. En effet, ces corpus auxquels seraient associés un ou plusieurs jeux de métadonnées, à définir, constituent une ressource extrêmement précieuse pour les chercheurs de toutes disciplines. Un des verrous actuel tient à la difficulté de connaître les corpus produits, les vocabulaires métiers utilisés, les référentiels scientifiques proposés, les conditions juridiques de la collecte des données, ce qui ne permet pas de réutiliser toujours les données collectées. Au-delà de ces questions, la prise en compte des nouvelles méthodes de structuration des données issues du *linked data* et du web de données permettront sans doute de mieux exploiter les corpus avec des outils de *search* (moteurs de recherche) ou encore de traitement de données. ISIDORE est déjà capable de moissonner du RDF (via le RDFa) et donc de relier les données entre elles pour mieux les exploiter. ISIDORE est donc déjà opérationnel pour le moissonnage massif des corpus. Il doit être l'un des vecteurs majeurs pour la diffusion et l'exploitation des corpus scientifiques proposés par les consortiums de la TGIR Corpus.

Il moissonne d'ailleurs un certain nombre de corpus de fonds d'archives proposés par les membres des consortiums de la TGIR Corpus (fabula.org ; CRDO-Aix ; Bibliothèques Virtuelles Humanistes, etc.) mais cela n'est possible que parce que les acteurs se connaissent et travaillent ensemble depuis plusieurs années. La structuration de la production des corpus permettra sans doute d'améliorer la circulation des bonnes pratiques qui sont déjà largement diffusées dans les communautés. Il s'agit d'accélérer maintenant les liens entre ISIDORE et les structures produisant des corpus, au travers de la TGIR Corpus par exemple, afin de placer cette réalisation comme le vecteur majeur de la diffusion des données de la recherche, archives et corpus de données inclus.

Exploiter les historiques des consultations pour proposer des recherches sur profils

Actuellement, un utilisateur qui utilise fréquemment la plate-forme ISIDORE doit

(ISIDORE collecte ; les portails tel que biblioshs font de la recherche fédérée en temps réel), mais d'imaginer des interopérabilités simples et « *user friendly* » pour les chercheurs.

²² Voir annexe 7.

²³ Pour *Application Programming Interface* ou interface de programmation, voir http://fr.wikipedia.org/wiki/Interface_de_programmation

renouveler sa ou ses requêtes et sélectionner les facettes qui l'intéressent à chaque consultation. On peut faire l'hypothèse qu'un chercheur qui travaille sur une thématique spécifique souhaiterait pouvoir conserver, pour une période déterminée, la formulation de ses requêtes et les facettes de sélection. Ceci implique de proposer aux utilisateurs la possibilité de demander l'enregistrement de ses choix afin de constituer un profil de recherche. Afin de respecter les données personnelles, un utilisateur pourrait à tout moment détruire définitivement son profil sans que les administrateurs d'ISIDORE puissent conserver ses données (mêmes anonymisées).

Dans le même ordre d'idée, les techniques actuelles de calcul de similarité permettent de proposer aux utilisateurs des choix d'articles (ceux qui ont consulté cet article ont aussi consulté ces articles). Cette fonctionnalité s'inscrirait dans la lignée du « Rebondir » actuellement proposé.

Proposer un service permettant de faire de la veille scientifique

Dans la filiation de proposer des recherches sur profils, l'une des voies de développement d'ISIDORE serait de proposer un service de veille scientifique utilisant l'API d'ISIDORE. Il s'agirait d'un « ISIDORE news » sur les données présentes dans ISIDORE qui pourrait être simplement incorporé dans les interfaces hommes/machines actuelles incluant, outre un profil utilisateur utilisant la fédération identité Education-Recherche de Renater²⁴ (qui est utilisée dans le monde universitaire français), la possibilité d'une identification via les réseaux sociaux (Facebook, LinkedIn) en particulier. En effet, l'utilisation massive des réseaux sociaux « grand public » par les étudiants pourrait être un vecteur important afin de placer ISIDORE dans la sphère des outils utilisés par les étudiants au quotidien, voir même les enseignants-chercheurs.

Par ailleurs, « ISIDORE news » est une demande exprimée par des utilisateurs d'ISIDORE rencontrés au fil des présentations et ateliers proposés en 2011.

Contribuer à la construction d'un vocabulaire scientifique pour le web de données

Le Web de données, dans lequel s'inscrit la plate-forme ISIDORE qui utilise les méthodes de structuration de l'information et des données (RDF, SKOS/RDF, OWL), les vocabulaires spécifiques (Dublin core element set, DC Terms, MODS, MIX, METS) et les mécanismes de référencement (URI utilisant par exemple des identifiants Handle ou Ark), a favorisé l'émergence de vocabulaires dédiés : FOAF pour *friends of a friends* par exemple qui permet de décrire des personnes, des objets, etc. ou encore SIOC pour les revues, ouvrages, blogs. De par sa capacité à pouvoir collecter des données issues du web de données, ISIDORE pourrait contribuer à faire adopter par la communauté scientifique des SHS les éléments fondamentaux des méthodes et techniques du web de données et de placer ainsi les données numériques de la recherche en SHS dans ce nouvel espace. Il est important que des outils tel qu'ISIDORE permettent aux communautés SHS de s'approprier le web de données afin d'en maîtriser, outre les enjeux techniques, les conséquences en matière d'édition de sources, de signalement (identifiants) et d'accès (y compris juridiques).

Améliorer la qualité des données moissonnées

Un des enjeux des années qui viennent est l'amélioration de la qualité des métadonnées et des données dont la production doit aujourd'hui se faire au plus près des acteurs (enseignants/chercheurs) dans les équipes de recherche. Cette idée est de plus en plus facilitée par les efforts déployés par la communauté des *digital humanities* qui organise régulièrement des formations, des séminaires et aménage des espaces d'échanges entre chercheurs, informaticiens, documentalistes, archivistes au cœur des équipes. C'est par l'échange quotidien entre les métiers que les conditions d'amélioration des métadonnées se font : par découverte et intérêt mutuel de

24 Voir : <https://federation.renater.fr> [d.c. 2011/11/12].

personnes impliquées sur un même projet de corpus par exemple.

Sur le plan technique cela veut dire aussi qu'il faut aujourd'hui dépasser, quand cela est possible, la surcouche applicative que forme, par exemple, OAI-PMH sur le web, afin de proposer des contenus structurés dans le web de données. Il convient de proposer des structururations riches à l'aide du RDF en ne se limitant plus au plus petit dénominateur commun qu'est le Dublin Core.

L'utilisation et l'élaboration de référentiels communs (thésaurus, lexiques, etc.), eux-mêmes proposés au travers du web de données, améliorera également la qualité des métadonnées. L'une des évolutions d'ISIDORE serait par exemple de proposer son ontologie et ses référentiels en SKOS/RDF en complément des référentiels tiers qui sont déjà dans le web de données (thésaurus W du SIAF, Rameau de la BNF, idref de l'ABES, etc.). Il s'agit donc de proposer, au travers d'actions concrètes et planifiées²⁵, la mise en place d'une politique d'amélioration qualitative des métadonnées entre les acteurs du domaine (TGIR par exemple).

Analyser les usages des étudiants

La formation à l'information scientifique et technique est l'une des préoccupations majeures du milieu universitaire ; il est reconnu que le niveau de maîtrise des outils de recherche documentaire par les étudiants a un fort impact sur la réussite de leur cursus universitaire.

De récentes études se sont penchées sur la question des pratiques informationnelles des étudiants dans le contexte universitaire. Ces enquêtes déplorent toutes un sérieux manque de maîtrise des outils de recherche documentaire. Le CREDOC concluait en 2005 à un déficit alarmant des compétences de recherche des étudiants de l'Université Pierre et Marie Curie (Paris 6) et de l'Université Denis Diderot (Paris 7)²⁶ et proposait de les combler en renforçant les efforts de formation.

En 2009, 65% des étudiants de 1ère année de LLCE de l'Université de Bourgogne²⁷ avaient encore le « réflexe Google » pour leur recherche d'information. Les services numériques de documentation de l'Université de Grenoble étaient reconnus comme incontournables par seulement 4% des étudiants²⁸ en 2007. En 2008, les archives ouvertes n'étaient rarement ou jamais utilisées par 77% des doctorants de l'Université de Bretagne²⁹. Force est de constater que, malgré les formations en recherche documentaire dispensées aux étudiants, il subsiste un décalage important entre l'offre électronique disponible et les pratiques effectives identifiées par les enquêtes menées dans le domaine universitaire.

Nous proposons d'aborder cette question selon une approche différente. En effet, mener une étude similaire évaluant les compétences informationnelles des étudiants nous mènerait probablement à des résultats semblables.

Étant donné que ces études montrent que les étudiants s'approprient peu les ressources documentaires universitaires, nous pouvons nous demander si ce ne sont pas les ressources documentaires qui devraient être adaptées aux habitudes informationnelles des étudiants. Une meilleure appréhension de leurs logiques d'usage passe par l'exploration de leurs pratiques informationnelles personnelles – et pas seulement universitaires. Cela suppose également d'envisager cette question selon le point de vue des étudiants et leurs besoins, et non selon le point

25 Les premières actions ont été déjà réalisées en ce sens par le pôle interopérabilité des données du TGE Adonis (<http://www.tge-adonis.fr/service/isidore>), dans le cadre de formations (ANGD d'Aussois par exemple) ou suite à des présentations dans les MSH (Lyon, Nanterre en particulier) ou des bibliothèques universitaires.

26 Maresca, Bruno, Dupuy, Claire, Cazenave, Aurélie, *Enquête sur les pratiques documentaires des étudiants, chercheurs et enseignants-chercheurs de l'Université Pierre et Marie Curie (Paris 6) et de l'Université Denis Diderot (Paris 7)*, 2005.

27 Martin, Aurélien, Perret, Cathy. *Les méthodes de recherche documentaire des étudiants de 1ère année de LLCE en début d'année*. Université de Bourgogne, CIPE, janvier 2009.

28 LIMA Laurent, FERNEX Alain, GRANON Arielle, Les étudiants grenoblois et Internet, Rapport de recherche, Enquête 2006-2007, Grenoble, Université Pierre-Mendès-France, 2007.

29 Henriot, Ottilia, Malingre, Marie-Laure, Serres, Alexandre, Enquête sur les besoins de formation des doctorants à la maîtrise de l'information scientifique dans les écoles doctorales de Bretagne (2008).

de vue des documentalistes ou des enseignants. Pour cela, il est important que cette étude soit réalisée par un ou une étudiante, voir un petit groupe d'étudiants.

Devant la montée en puissance des pratiques culturelles numériques, de l'essor des forums, des blogs et des réseaux sociaux, on ne peut imaginer des étudiants dépourvus de toute compétence en matière de recherche informationnelle. Notre première hypothèse est qu'il y a un écart entre leurs pratiques informationnelles au niveau personnel et leurs pratiques au niveau universitaire. Il convient donc d'explorer leurs habitudes informationnelles dans leur vie privée (même si la frontière entre pratiques personnelles et pratiques universitaires va s'avérer très probablement perméable). Outre la recherche documentaire via le couple Google/Wikipédia, les pratiques phares de l'utilisation de l'Internet sont la consultation de boîtes mails, l'utilisation des outils de messagerie instantanée et la création, diffusion ou critique de contenus. Les pratiques numériques révèlent le besoin de disposer d'un maximum d'informations en un minimum de temps. Face au succès des plate-formes de formation « Moodle », qui combinent l'accès à des ressources numériques et l'incorporation d'outils de participation, de collaboration et de discussions inspirés des pratiques personnelles des étudiants, notre seconde hypothèse est que la fréquentation et les compétences des étudiants, en matière de recherche informationnelles, peuvent évoluer grâce à l'incorporation d'outils qu'ils utilisent habituellement à titre personnel.

L'étude comporterait trois étapes :

- a) obtenir des indications sur les pratiques informationnelles personnelles ;
- b) dresser l'état des lieux pratiques universitaires en matière de recherche d'information ;
- c) évaluer la pratique d'un outil proposant des activités communément utilisées à titre personnel.

Ces trois types de données nous permettront d'une part d'appréhender s'il y a corrélation entre pratiques personnelles et universitaires et, d'autre part, de mesurer l'impact de l'intégration d'outils tels que la messagerie instantanée, le forum, les wiki, sur la fréquentation d'un site de ressources informationnelles.

Intégrer des données scientifiques multilingues

Actuellement, ISIDORE moissonne des publications scientifiques rédigées en d'autres langues que le français. En décembre 2011, sur les 1 393 746 ressources moissonnées, 23 langues étaient représentées, la deuxième langue étant l'anglais avec 72 986 ressources. Les interfaces d'interrogation sont en français mais il est tout à fait possible de taper une requête dans une des 23 langues voir même de sélectionner par langue. Néanmoins, les techniques d'indexation d'ISIDORE sont fondées sur les thésaurus monolingues et de plus les expansions sont fondées sur la langue française ce qui a conduit à limiter l'utilisation de Geonames.org pour éviter des conflits entre des noms d'origine de ville étrangères et des noms français. Il en est de même pour les fonctionnalités de « Rebondir ».

La prise en compte des ressources multilingues pose des problèmes complexes et nécessite l'exploitation de thésaurus multilingues alignés entre eux. Néanmoins, certaines disciplines ne publient pratiquement plus qu'en anglais (par exemple la biologie, l'informatique, etc.), et si les disciplines de l'INSHS sont moins touchées par ce phénomène, c'est néanmoins déjà le cas de certaines sous-disciplines de la psychologie, de l'économie, de la linguistique, etc.


Il semble donc opportun de travailler à l'intégration de données multilingues en réalisant un premier inventaire des thésaurus disponibles et des besoins en terme d'interrogation. Ce travail exploratoire devra déboucher sur une réflexion importante dont il faudra bien mesurer les trajectoires. La complexité d'une indexation multilingue sur plus d'un 1 000 000 de ressources nécessitera sans doute une profonde refonte du cœur d'ISIDORE (chaîne de traitement ; outils de gestion de référentiel ; IHM).

Références

- Becker, H., 2009, *How to Find Out How to Do Qualitative Research*, International Journal of Communication 3, 545-553.
- Bo V., 2007, *Scientific publishing: some food for thought*, Journal of Science Communication, vol. 6, n° 1, 2007.
- Dacos M., Mounier P., 2011, *L'édition électronique*, Editions La Découverte, Paris.
- Dargentas M., Brugidou M., Le Roux D., Salomon A.-C. (dir.), à paraître, *L'analyse secondaire en recherche qualitative : une nouvelle pratique en sciences humaines et sociales*, Paris, Lavoisier.
- Filliatreau, G., 2008, *Bibliométrie et évaluation en sciences humaines et sociales : une brève introduction*, Revue d'histoire moderne et contemporaine, 55/4bis, 61-66. Genèses, « Sciences sociales : archives de la recherche », n°63.
- Ginsparg P., 1996, *Electronic Publishing in science – winners and losers in the global research village*, Joint ICSU Press/UNESCO Expert Conference on Electronic Publishing, UNESCO, Paris, en ligne : <http://www.library.illinois.edu/icsu/ginsparg.htm>
- Maurel L., 2008, *Le défi du droit d'auteur*, Presses de l'ENSSIB.

Annexe 1

Copie écran de la une d'ISIDORE le 18 décembre 2011



Vous êtes ici : [Accueil](#)

Votre recherche

Sciences
Humaines et Sociales

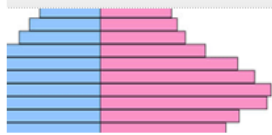
> CHIFFRES

Collections	58
Sources	1 255
Ressources	1 393 746

> NAVIGUER

- > Par siècles
- > Par types de documents
- > Par catégories
- > Par périodes historiques
- > Par disciplines
- > Par langues
- > Par nature des documents
- > Par nature des sources
- > Par collections et organisations

> AUJOURD'HUI ZOOM SUR...



Démographie



Yacuiba: un islote de la mundializacion en el Chaco boliviano
Souchaud, Sylvain *et al.* (2007)

Se trata de un capítulo sobre las migraciones humanas en la ciudad de Yacuiba (Sur de Bolivia, frontera con Argentina), que forma parte del libro: Movilidad y procesos migratorios en el espacio de frontera argentino-boliviana. Este libro surge de la investigación de los procesos migratorios que no solo ...


Source : HALSHS

[Voir plus d'informations](#) | [Voir la fiche de la ressource](#)

> CONTRIBUER

Bibliothèques, centres de documentation, laboratoires et projets de recherche, vous pouvez [contribuer à rechercheisidore.fr](#) avec vos données.


> S'INFORMER

 **2e Appel d'offres conjoint STDF - IRD**
18 déc. 2011
L'Institut de recherche pour le développement (IRD) et le Science and Technology Development Fund (STDF) égyptien annoncent le lancement de leur deuxième appel d'offres conjoint. Avec un budget de 1,8 millions d'euros, il a pour objectif

> SOURCES

Annexe 2

Un exemple de réponses à la requête “Culture numérique” le 18 décembre 2011



Vous êtes ici : [Accueil](#) > [Recherche](#)

> RECHERCHER

[Plus d'options](#)

> AFFINER

> Par siècles


- ☐ 1300 (4)
- ☐ 1400 (4)
- ☐ 1500 (6)
- ☐ 1600 (12)
- ☐ 1700 (86)
- ☐ 1800 (212)
- ☐ 1900 (677)
- ☐ 2000 (5176)

> Par types de documents

- ☐ Articles (1443)
- ☐ Autres (134)
- ☐ Autres (622)
- ☐ Billets (242)

> RESULTATS DE LA RECHERCHE

> **6109 résultats** (318 msec) > Par 10 > Vue Liste




Journée d'étude: Culture numérique: Nouveaux espaces d'expression et de création adolescentes (jeu vidéo, blog, téléphone portable.)

Mathilde Lévêque (15 sept. 2009)

Cette journée d'étude organisée par Lecture Jeunesse aura lieu le 15 octobre 2009 à la Cité des sciences et de l'industrie 30 avenue Corentin Cariou, 75019 Paris, métro Porte de la Villette (ligne 7) Les outils numériques de communication et de divertissement offrent à leurs utilisateurs des possibilités ...

[Voir plus d'informations](#) | [Voir la fiche de la ressource](#)




Synthèse du Programme de Numérisation pour l'Enseignement et la Recherche (PNER, 1999-2001)

Noyer, Jean-Max (2001)

Synthèse des travaux du Programme PNER (1999-2001) mis en oeuvre par la Maison des Sciences de l'Homme et soutenu par le Ministère de l'Education Nationale, de la Recherche et de la Technologie. Ce programme a analysé les évolutions des usages, proposé une réflexion sur les normes, travaillé la question ...

[Voir plus d'informations](#) | [Voir la fiche de la ressource](#)



Métier d'élève, métier d'enseignant à l'ère numérique

Dioni, Christine (28 févr. 2008)

Annexe 3

Un exemple d'une ressource enrichie



Vous êtes ici : [Accueil](#) > [Recherche](#) > [Fiche ressource](#)

> FICHE DE LA RESSOURCE



Journée d'étude: Culture numérique: Nouveaux espaces d'expression et de création adolescentes (jeu vidéo, blog, téléphone portable.)

Par : [Mathilde Lévêque](#)

Date : 15 septembre 2009 | disponible sur <http://magasindesenfants.hypotheses.org/378>

Cette journée d'étude organisée par Lecture Jeunesse aura lieu le 15 octobre 2009 à la Cité des sciences et de l'industrie 30 avenue Corentin Cariou, 75019 Paris, métro Porte de la Villette (ligne 7) Les outils numériques de communication et de divertissement offrent à leurs utilisateurs des possibilités technologiques importantes. Les jeunes sont parmi les [...]

> APERÇU DE LA RESSOURCE

> REBONDIR ?

> ENRICHISSEMENTS

Culture Culture Expression Création
Adolescentes Jeu jeu Vidéo Téléphone
Lecture lecture jeunesse Jeunesse lieu
Octobre cité-ville Classification internationale
type de l'éducation sciences Sciences
artisanat Industrie Pari porte Outils
Communication Jeunes Congrès et
conférences Congrès

☒ Rameau ☒ Pactols


> ESPACES GÉOGRAPHIQUES

> PARTAGER

Partager cette ressource

Annexe 4

L'interface du Sparql endpoint



Vous êtes ici : [Accueil](#) > [SPARQL endpoint](#)

> **SPARQL endpoint**

Entrez votre requête SPARQL :

```
SELECT DISTINCT ?Concept WHERE {
  [] a ?Concept.
}
```

Format de sortie des résultats :

> **Qu'est ce qu'un SPARQL endpoint ?**

Un *SPARQL endpoint* est une interface web qui permet d'interroger de l'information numérique structurée en *RDF* (*Resource Description Framework*). L'interrogation se fait à l'aide du langage *SPARQL*, langage normalisé et ouvert, développé et maintenu par le *W3C*.

> **Avec le SPARQL endpoint d'ISIDORE, à quoi a-t-on accès ?**

Le *SPARQL endpoint* d'ISIDORE permet d'interroger l'ensemble des métadonnées moissonnées, structurées et enrichies par ISIDORE. C'est à dire le même contenu que celui qui est accessible au travers du site web [rechercheisidore.fr](#). Les métadonnées moissonnées sont enrichies par traitements sémantiques à l'aide de référentiels scientifiques (thésaurus, listes d'autorités, vocabulaires). Actuellement, l'ensemble des enrichissements est accessible via le *SPARQL endpoint*.

> **Du SPARQL et du RDF pourquoi faire ?**


Avec RDF et le langage SPARQL il est possible d'interroger l'information structurée contenue dans les métadonnées sans la limite du plus petit dénominateur commun. Surtout il est possible, puisqu'il existe de nombreux entrepôts de données structurés selon RDF, de construire des applications web ou mobiles qui vont utiliser ces données RDF en utilisant plus particulièrement le fait qu'elles sont reliées entre elles par des URI qui prennent la plupart du temps la forme d'URL, c'est à dire des adresses web. C'est le principe du *linked data*. Pour illustrer ce principe nous vous proposons une application web, le *ISIDORE terms navigator simple edition*, qui est un tableau de bord sémantique qui fonctionne à l'aide du *SPARQL endpoint*.

> **Pourquoi du RDF ?**

Mis au point par le W3C dans le cadre des activités du Web sémantique, RDF n'est pas à proprement parler un schéma de métadonnées. Il constitue un **modèle de description des données structurées** inspiré de la logique des prédicats de premier ordre et de la théorie des graphes. Sa généralité et sa souplesse offrent un cadre d'interopérabilité pour la description de tous types de ressources dans un environnement en réseau comme le Web. RDF est un modèle qui permet d'exprimer des assertions selon un modèle très simple comparable à une phrase simple : [sujet] [prédicat] [objet]. Chaque assertion forme un triplet dont les différents composants sont exprimés sous la forme d'une URI. L'intérêt de RDF réside dans le fait qu'il est possible d'exploiter des triplets RDF quelque soit le vocabulaire utilisé sans conversion, à l'inverse de XML pour lequel il est nécessaire de convertir les données et elles structurées selon le même schéma. Ainsi il est possible de combiner différents ensembles de données d'origine existante pour en tirer de nouvelles informations.

Annexe 5

Un accès au catalogue du Sudoc par le connecteur synchrone SRZ (protocole SRU/SRW)



Vous êtes ici : Accueil > Recherche > Recherche bibliographique

> A PROPOS

La recherche bibliographique d'ISIDORE vous permet d'interroger, par un simple clic, des catalogues de bibliothèques. Votre recherche initiale est ainsi portée vers plusieurs catalogues bibliographiques.

> BIBLIOTHEQUES

☒ SUDOC (503)

> RECHERCHE BIBLIOGRAPHIQUE

> 503 résultats pour les 1 bibliothèques

Mot-clé : culture numérique

SUDOC (503 résultats) > Par 5

[une nouvelle pédagogie pour une culture de l'information à l'heure du numérique](#)
dc:type : Texte imprimé
dc:publisher : [Limoges]
dc:date : impr. 2011
dc:source : 1 vol. (159 p.)
dc:relation : 978-2-916571-54-6 (ISBN ,)


[Nouveaux accès, nouveaux usages à l'ère numérique, la culture pour chacun des actes du forum d'Avignon culture, économie, médias, 4-6 novembre 2010 / acts of the forum d'Avignon culture, économie, médias, 4th-6th november 2010](#)
dc:type : Texte imprimé
dc:publisher : Paris
dc:date : impr. 2011
dc:source : 1 vol. (538 p.- 32. p. pl.) ;
dc:relation : 978-2-07-013526-4 (ISBN ,) , 2-07-013526-8 (ISBN ,)

[Hommage à Henri Bartoli](#)
dc:type : Texte imprimé
dc:publisher : Paris
dc:date : DL 2011
dc:source : 1 vol. (106 p.)
dc:relation : 978-2-296-56533-3 (ISBN ,)


[Induced transmembrane voltage and electroporation of cells in cultures in vitro](#)
dc:type : Ressource électronique


Annexe 6

Exemple d'intégration ISIDORE : Catalogue local de l'Université de Bordeaux 3 (SHS) BABORD+ (onglet "rebondir"). ISIDORE est positionné au même plan que Gallica, SUDOC.



L'interface Babord+ a ouvert le 1er septembre. Cette interface est pour le moment disponible en version bêta et certains bugs peuvent subsister. Le temps de les corriger, l'ancienne version du catalogue reste disponible.


[Contact](#) | [Mentions légales](#) | [Aide](#)

BABORD  Version bêta

Accueil > Résultats

[Affiner ma recherche](#)
[Résultats](#)
[Autre recherche](#)

CRITERES CHOISIS

☒ Tous : djibouti


Nombre de résultats par page : 10
 Tri : Pertinence
 Décroissant

Page 1 de 13

1/130

Djibouti : République de Djibouti


Otayek, René
Bruylant 1987

 Livre

2/130

Dossier national : Djibouti

Dini, Abdallah
Centre international de formation et de recherche en population et développement 1995

 Livre

Rebondir

[Sudoc](#) (catalogue des universités françaises)
[Factiva](#) (presse)
[Gallica](#) (Bibliothèque numérique de la BnF)
[ISIDORE](#) (moteur de recherche en sciences humaines et sociales)
[BnsA](#) (Banque numérique du savoir d'Aquitaine)

Nuage de mots

Nouvelle recherche

Annexe 7

Intégration d'ISIDORE dans l'outil de recherche de Microsoft Window 7

